An attempt to validate national mean scores of Conscientiousness: No necessarily paradoxical findings

Running head: VALIDITY OF CULTURE-LEVEL PERSONALITY SCORES

An Attempt to Validate National Mean Scores of Conscientiousness: No Necessarily Paradoxical Findings

René Mõttus [1, 2]

Jüri Allik [1]

Anu Realo [1]

[1] Department of Psychology, University of Tartu, The Estonian Center of Behavioral and Health Sciences

[2] University of Edinburgh Center of Cognitive Ageing and Cognitive Epidemiology, UK

Mailing Address: Dr. René Mõttus
Department of Psychology
University of Edinburgh
7 Georges Square, EH8 9JZ, Edinburgh
United Kingdom

E-mail: rene.mottus@ut.ee

Phone: +372 7375902

Abstract

Three large cross-cultural databases on personality were used to study the relationship between culture-level mean scores of Conscientiousness and 18 country-level criterion variables including health indicators, religiosity, democracy, corruption, economic wealth and freedom, and the presence of a favorable business environment. Mean Conscientiousness scores were significantly related to most of the criteria but different facets of Conscientiousness had very different relationships with external criteria. In several facets, the patterns of relationships to the external criteria were consistent with clearly formulated predictions. The pattern of relationships was moderated by the type of ratings (self- vs observer-ratings). More rigorous requirements for the study of the culture-level relationships between personality and criterion variables are discussed.

An Attempt to Validate National Mean Scores of Conscientiousness

National mean scores of personality traits often demonstrate a consistent pattern in their geographic distribution (Allik & McCrae, 2004; Schmitt, Allik, McCrae, & Benet-Martinez, 2007), a substantial level of agreement between self- and other-ratings (McCrae, Terracciano, & 79 Members of the Personality Profiles of Cultures Project, 2005), and a meaningful configuration of correlations with several relevant culture-level indicators (Allik & McCrae, 2004; Hofstede & McCrae, 2004; McCrae, Terracciano, Realo, & Allik, 2008). Recently, however, the validity of national mean scores of self- and other-reported personality traits has been seriously questioned, at least in the Conscientiousness domain. Heine and colleagues (2008) reanalyzed published data and showed that aggregate national scores of self-reported Conscientiousness were, contrary to the authors' expectations, negatively correlated with various country-level behavioral and demographic indicators of Conscientiousness, such as postal workers' speed, accuracy of clocks in public banks, accumulated economic wealth, and life expectancy at birth. Oishi and Roth (2009) expanded the list of contradictory findings by demonstrating that nations with high self-reported Conscientiousness were not less but more corrupt.

These and other similar warnings against taking national means of self- and peer-reported personality at face value (Ashton, 2007; McCrae, Terracciano, Realo, & Allik, 2007; Perugini & Richetin, 2007) must be heeded. However, for time being there is still too little in the way of comprehensive research on the ability of culture-level personality scores to predict relevant external and, preferably, objective criteria. Before we can come to a final verdict on the predictive validity of culture-level personality scores, we need further research that is both more theoretically driven and methodologically comprehensive. We believe that the existing research should be extended in two principal ways.

Firstly, it is useful to look at personality traits in a more differentiated manner. So far, validity studies have used broad personality traits such as the Big Five domains (Heine et al., 2008; Oishi & Roth, 2009), but these broad traits can be broken down into more specific lower-level facets. A more differentiated description of personality traits is justified by the possibility that external criterion variables may not so much be correlated with the central core of the trait (i.e., the common variance of different lower-level facets of the trait) but, rather, with some more peripheral aspect of it. At the level of individuals, different facets of Conscientiousness have been found to relate differently to external criteria (Roberts, Chernyshenko, Stark, & Goldberg, 2005), with the correlations sometimes even having different signs (Moon, 2001). Distinguishing between the different facets of Conscientiousness may render the previous unexpected relationships more easily understandable.

To give an example, dictionaries usually define an extrovert as a "person primarily interested in the people and things around him rather than in his inner thoughts" (Williams, 1979). Taking this definition literally, it would be hard to believe that Norwegians, Danes, and Swedes have the highest—and Italians, Portuguese, and Russians relatively much lower—scores on the NEO PI-R self-rated Extroversion scales (McCrae, 2002). However, these results become less perplexing once we come to understand that "turning attention outside" is not the main core of the scientific definition of Extroversion. Although sociability is undoubtedly an important part of Extroversion, extroverts' sociability may be a by-product of the core of Extroversion (e.g., reward sensitivity), rather than the core feature itself (Lucas, Diener, Grob, Suh, & Shao, 2000). If positive emotions and life satisfaction form the true core of Extroversion, the ranking of countries presented above becomes more consistent with our intuition. Similarly, a negative correlation between a seemingly obvious criterion variable

and culture-level Conscientiousness may be driven by only one facet of Conscientiousness (e.g., traditionalism) which, arguably, is not directly related to the core of Conscientiousness.

Secondly, going for more refined description of personality should be coupled with more rigorous and comprehensive choice of criteria. The results of predictive validity studies obviously depend heavily on the external criteria used. Ideally, criteria choice should be based on a clear, theoretically sound account of the causal chain of events that connect the ways of responding on personality scales to variation in the expected external criterion variable. The links are, however, not always as transparent as they have been assumed to be. For instance, it has been tempting to hypothesize that high culture-level means of Conscientiousness should yield high accuracy of bank-clocks but, in fact, it needs a causal explanation *how* a greater proportion of conscientious people in a given population helps to get bank clocks more accurate. To make the matters even more complicated, bank clocks are monitored by a very small, perhaps seriously unrepresentative minorities of populations. Based on these issues, there are at least two questions to ask when choosing any validity criteria for culture-level personality scores.

First of all, is there certainly only one possible way how personality trait scores can relate to the criterion? This may seem as a trivial questions but its answer is crucial. Let us consider the example of a hypothetical relationship between Conscientiousness and democracy. Democracy provides political and civil rights, allowing people to have freedom of choice in their public and private actions (Inglehart & Welzel, 2005). It could be argued that the maintenance of democracy presupposes not only efficient regulation and a transparent legal system but also competent and responsible people. Therefore, one could expect that in more democratic countries citizens are more responsible and disciplined, resulting in a positive correlation between the level of democracy and mean national scores of Conscientiousness.

However, the relationship may almost equally well go the other way around—it is dictatorship that better enforces hard work, discipline, and order in society. According to Inglehart (Inglehart & Welzel, 2005), effective democracy is much more likely to be found in cultures with a strong emphasis on self-expression values, whereas dutifulness, order, and hard work are the correlates of survival, the opposite of self-expression. Thus, it can be argued that in countries with higher scores of Conscientiousness—that is, where people are rule-abiding, inhibited by social constraints, and keen on keeping order—people are not able to realize their potential for freedom and autonomy which, in turn, are the cornerstones of democracy. So, in principle, the relationship between democracy and Conscientiousness at the national level could be either positive or negative. Alternatively, the relationship depends on the facet of Conscientiousness that we are looking at: self-perceived competence may be positively related to the possibility of having political say, whereas more autocratic/totalitarian societies boost higher levels of orderliness, hard work and cautiousness.

Next, mean personality scores are supposed to describe societies in general. The same has to be asked about the criteria—in which way they describe societies? Do they describe societies in a general manner without referring to any specific individuals (e.g. the level of democracy), or do they reflect some sort of aggregate statistics that result from the behavior of many specific individuals (e.g. percentage of smokers), or do they refer to the behavior of only a very tiny fraction of society (e.g. rate of rare diseases)? It may be that the relationships are straightforward: these are the “average” people who mostly contribute to the general societal characteristics such as democracy, and the average personality scores are reflected in the behavior of each society member in a similar way. But this is not necessarily always the case. For example, nobody doubts that people commit suicide mainly because they feel desperately unhappy. Indeed, the level of an individual’s life dissatisfaction (i.e., depression

and negative emotions) is a strong predictor of suicide intention (Koivumaa-Honkanen et al., 2001). However, at the aggregate national level, there is a strong positive association between happiness and the suicide rate: in countries where people are generally happy and satisfied with their lives, the suicide rate is higher than in those countries where people tend to feel more miserable (Bray & Gunnell, 2006; Diener & Diener, 1995). An explanation for this paradox is that the very small number of people who commit suicide may be mainly those who are not able to cope with the social demand for being happy brought about by the relatively high average level of happiness (Inglehart, 1990). As another relevant example, antisocial behavior has typically been related to low Conscientiousness in individuals (Miller & Lynam, 2001) whereas at the level of nations this may be the other way around. Societies with generally less conscientious people may, as a compensatory mechanism, develop stricter rules to cope with crime (e.g. via religion), resulting in that the few individuals inclined to crime (e.g. murder) are better constrained. Thus, statistics that reflect the behavior of only a fraction of people may not always be straightforward criteria for average personality scores. Likewise, many general characteristics of societies such as democracy or economic freedom may be shaped by a small minority of people, possibly obscuring the relationships between average personality and nation-level outcomes.

How Should External Validity Criteria for Conscientiousness be Chosen?

Given that we hardly have any solid theories on which to base our selection of "objective" external criteria for aggregate culture-level personality scores, the most straightforward and transparent strategy would then be to extrapolate from individual-level findings. For instance, we know that more conscientious individuals are more likely than less conscientious people to start their own businesses (Zhao & Seibert, 2006) and we may expect that a similar tendency is also true for culture-level analyses: that in countries with a higher

concentration of conscientious people, entrepreneurship is encouraged. At the same time, as described above, there are several potential (related) issues related to choosing appropriate criteria for culture-level personality scores: the criteria may not be related to all aspects of the trait in the same manner, there may sometimes be equally tenable alternative ways for personality scores to relate to criteria, and the criteria may not describe the society in general.

In such a situation, a viable approach is to use multiple criteria. Normally we do not rely on one single participant to demonstrate a relationship—there are too many chances to be wrong. We sample numerous individuals, which should allow the hypnotized relationship to shine through the "noise" and reduce the chances of being wrong. The same is true for the predictor-criterion relationships. Due to the theoretical immaturity of the personality-culture interface, a single validation criterion might well be unrelated to personality traits or the relationship may even run in the opposite direction to the reasonable expectations of the researcher. For multiple criteria this is less likely. If aggregate personality traits are valid operationalizations of culture-level personality, patterns of meaningful relationships should be expected, even if a number of specific predictor-criterion relationships seem to be controversial. Furthermore, in a pattern of associations, even seemingly paradoxical relationship may finally make sense.

Aims of the Study and the Rationale behind the Selection of External Validity Criteria

In this study, we analyze how different aggregate ratings of Conscientiousness—both self- and other-reports—are related to a diverse set of potential criterion variables. Improving on previous studies (Heine et al., 2008; Oishi & Roth, 2009), we investigate the relationship at the facet scale level of Conscientiousness and employ a sample of different types of external validity criteria.

To make full use of the facet level approach, we set up hypotheses separately for each facet scale. Since we believe that criteria might relate to different facets of Conscientiousness in a different manner, setting up precise hypotheses only for Conscientiousness domain in general would not be reasonable. In addition, in order to quantify the match between hypothesis and the actual findings based on available data, we set up hypotheses in terms of predicted correlations (see Table 1). That is, for each criterion its hypothesized correlations with all domain and facet scales were listed and later compared to the respective empirical correlations. We believe that in the situation where we have the multi-trait multi-method data predicting multiple variables, quantifying the patterns of relationships is the most straightforward way of getting from numbers to conclusions (cf. Westen & Rosenthal, 2003). We realize that there is a great degree of arbitrariness in the predicted correlations but this is the only transparent way to organize and test the complex pattern of hypotheses. Since Conscientiousness domain scores consist of the contributions of various facets, we will not set up point-predictions for the domain scores.

Insert Table 1 about here

We employ potential criteria from all of the three aforedescribed categories: (1) those that reflect behaviour and outcomes of many individual culture-members (e.g., prevalence rate of some common diseases or health-related behaviors), (2) those that reflect the behavior and outcomes of a few culture-members, and (3) those that reflect the functioning of societies as a whole (e.g., democracy). Such broad sampling of validity criteria should reduce the risk of ending up with incorrect conclusions due to uninformed expectations. In the first category ("broadly representative individual-level indicators"), for each individual the probability of

contributing to the statistics is relatively high, which means that extrapolating from individual-level relationships to culture-level prediction is the most transparent. In the second category ("unrepresentative individual-level indicators" such as prevalence of HIV or homicide rate), only a few people contribute to the criteria in most countries but we cannot rule out a priori that the number of those a few people in some way reflects the general tendencies in the societies. In the third category ("societal-level indicators"), the relationships of the criteria to personality scores seem conceivable (as discussed below) but the mechanics are to a large extent based on theoretical guessing and, as a result, are less transparent.

Religiousity. In the category of the so-called broadly representative individual-level indicators, religiousness is the first external validity criteria. In numerous individual-level studies, religiousness has been shown to be related to Conscientiousness: individuals who admit that religion plays an important role in their lives, in general, score higher on Conscientiousness (Lodi-Smith & Roberts, 2007; Saroglou, 2010). Religion often emphasizes needs for self-control, persistence, order, work (especially Protestantism) and a highly organized life. Sargolou (2010) concluded in his meta-analysis that the relationship is substantive: religiousness seems to be the result rather than the cause or a covariate of high Conscientiousness. As a result, we expect that the prevalence of religiousness in cultures relates positively to four culture-level mean personality scores: most strongly to Self-Discipline, Achievement Striving and less to Dutifulness and Order. High competence, on the other hand, should characterize secular cultures that stress ability and freedom of individuals to manage their lives, while in more religious cultures power is believed to be the domain of God. We do not believe that religiousness should necessarily be related to Deliberation.

Life expectancy and health behavior. At the level of individuals, Conscientiousness is positively related to life expectancy (Deary, Batty, Pattie, & Gale, 2008; Friedman, Tucker,

Schwartz, Tomlinson-Keasey, & Martin, 1995; Kern & Friedman, 2008). An obvious explanation for the link is that more conscientious people live healthier lives and are less likely to be engaged in health-risk behaviors. Bogg and Roberts (2004) reviewed the literature on the relationship between Conscientiousness and health-related behaviors and concluded that less conscientious people, indeed, tend to overuse alcohol, eat unhealthy food, smoke and use drugs, and engage in unsafe sex, risky driving, and violence. Low Conscientiousness also predicts obesity (Terracciano et al., 2009). Although the effect sizes are usually not large, the overall pattern tends to be consistent—low Conscientiousness is related to numerous aspects of an unhealthy lifestyle (Ozer & Benet-Martinez, 2006). Extrapolating from these findings, we expect that several facets of culture-level aggregate Conscientiousness scores are related to several health-related outcomes that characterize a large proportion of people such as prevalence rates of alcohol abuse, tobacco use, obesity and death from cardio-vascular disease. As well, we expect the facets of Conscientiousness to relate to less common outcomes such as the prevalence of drug-use, the prevalence of sexually transmitted diseases such as HIV, the prevalence of death related to motor vehicle accidents, injuries and homicides. In particular, we assume that the prevalence of unhealthy behaviors and resulting diseases is most strongly related to low Self-Discipline and Deliberation and to somewhat lower but constant extent to other four facets.

Democracy. In the category of societal-level indicators, we start with democracy. As mentioned earlier, democracy characterizes cultures with higher self-expression and low survival-oriented values. As a result, we believe that democracy is negatively related to three Conscientiousness facets: most strongly to Deliberation (being cautious is perhaps very useful for doing well in totalitarian/autocratic societies), and to lower degree to Order and Achievement Striving. At the same time, it is likely that democracy is related to higher

perceived mastery (i.e. high Competence). High Dutifulness and Self-Discipline may characterize both democratic and non-democratic societies.

Economic freedom and business environment. Next, economic freedom and a smoothly functioning business environment should reflect highly Dutiful and Orderly administrators. Economic freedom should also be positively related to Competence and Achievement Striving as it plays on individuals' abilities and ambitions to achieve economic success. Societies with overly cautious people (high Deliberation), on the other hand, may want to refrain from giving people much freedoms and responsibilities. The relationship of Self-Discipline to economic freedom is difficult to predict, thus we hypothesize that there is no association. Among other things, economic freedom should reflect the ease of starting new businesses. In a free business environment it should take shorter time than in less free societies, to start a new business. Therefore, we expect the amount of time it takes to start a new business to co-vary with the facets of Conscientiousness in an exactly opposite manner to economic freedom. Similarly, border delays and complicated shipping of goods provide tests of Orderliness and Dutifulness of administrators while high Cautiousness, on the other hand, may lead to more complicated shipping procedures. Thus, these criteria should relate to facets of Conscientiousness in the same way as the number of days to start business. Corruption should be most strongly related to low Dutifulness and Cautiousness. Refraining from the temptations should also be more likely in case of high Self-Discipline. It is more difficult to see its relationships with Competence, Order and Achievement Striving, leading us to predict no correlation.

Human development. As a general indicator of a well functioning society, we look at the level of human development (as reflected by the Human Development Index; HDI). We expect that high Competence is the strongest predictor of high HDI, while Order, Dutifulness,

Achievement Striving and Self-Discipline have somewhat lower positive correlations with the HDI. Being too Cautious (high Deliberation), on the other hand, may at times hinder development. As a result, we predict that Cautiousness will have no relationships with the HDI. We predict the same pattern for GDP, another rough indicator of development. Finally, with respect to life-expectancy, Cautiousness is also likely to be helpful on top of other Conscientiousness facets.

Wealth of nations. Cross-cultural researchers tend to believe that culture-level relationships should be analyzed free of economic effects (e.g., Hofstede, 2001). Indeed, it is reasonable to expect that much of the variance in health statistics, HDI and other selected criteria is explained by economic wealth that, in turn, tends to relate to aggregate culture-level Conscientiousness (Heine et al., 2008). Thus, it is possible that many of the predictor-criterion relationships are just spurious correlations reflecting the operation of national wealth-Conscientiousness associations. Therefore, we also investigated the effect of controlling for Gross Domestic Product (GDP) per capita of the nations in question.

Finally, we aimed to use only those criteria for which relationships with aggregate Conscientiousness could be calculated in at least 30 countries. In fact, even this sample size is suboptimal because only correlations bigger than $r = .46$ are statistically significant at $p < .01$ (in order to reduce the probability of Type I errors resulting from multiple comparisons we chose a 1% alpha level). This effect size is rather remarkable as, for instance, some of the most solid validity coefficients in psychology, such as those of general mental ability in predicting school and professional achievements, tend to be in the same range (Strenze, 2007). However, we realized that setting a higher minimal number of cultures for each relationship may seriously cut down the number of available criteria.

We also report some the correlations of Conscientiousness scores with GDP, life-expectancy and corruption although these relationships have been reported elsewhere (Heine et al., 2008; Oishi & Roth, 2009). The reasons for this are that in the present study we make use of facets scale scores in addition to broad domain scores and we have personality scores for a somewhat bigger sample of culture than has previously been available.

Method

Personality Measures

Self-reported NEO PI-R measures. Aggregate self-rated NEO PI-R Conscientiousness scores (for 6 facet scales and their average) were taken from McCrae (2002), who reported the data for 36 countries. Additionally, mean self-reported NEO PI-R scores for three African nations not represented in McCrae's (2002) data were provided by J. Rossier (personal communication, October 28, 2008): Mauritius, Republic of Congo, and Democratic Republic of Congo. For Lithuania and Poland, the mean scores were obtained from studies by Zukauskiene and Barkauskiene (2006) and Siuta (2006), respectively. For Finland, the data collected by Lönnqvist and colleagues (2007) were used. National mean scores had been or were transformed into *T*-scores using the U.S. norms given by Costa & McCrae (1992). Thus, in total, aggregate self-report NEO PI-R scores were available for 42 nations.

Observer-reported NEO PI-R measures. Aggregate scores for observer-reported NEO PI-R Conscientiousness scores were collected by the members of the Personality Profiles of Culture Project (McCrae, Terracciano, & 78 Members of the Personality Profiles of Cultures Project, 2005). The mean scores of 51 cultures, standardized and transformed into *T*-scores relative to international means, are reported in McCrae and Terracciano (2008).

External Validity Criterion Variables

Atheism. This variable relates to the percentage of people saying that they are atheist, nonbelievers, or not believers in a "personal" God (Lynn, Harvey, & Nyborg, 2009; Zuckerman, 2007). This is the inverse measure to culture-level religiousness.

Health statistics. We chose the following variables from the database of the World Health Organization (2009) (the latest year with available data was used): (1) Alcohol consumption among adults aged ≥ 15 years (liters per person per year; 2003); (2) Prevalence of current tobacco use among adults aged ≥ 15 years (%; 2005); (3) Prevalence of obesity (Body Mass Index ≥ 30) among men and women aged ≥ 15 years (%; 2000 – 2007); (4) Prevalence of HIV among adults aged ≥ 15 years (per 100 000 population; 2007); (5) Age-standardized mortality rates by cardio-vascular diseases (per 100 000 population); and (6) Age-standardized mortality rates by injuries (per 100 000 population).

Traffic-related deaths. The annual number of road fatalities per capita in each country were taken from http://en.wikipedia.org/wiki/List_of_countries_by_traffic-related_death_rate on June 29, 2009

Homicide rate. The annual number of homicides per 100 000 population in each country were taken from http://en.wikipedia.org/wiki/List_of_countries_by_homicide_rate on June 29, 2009.

Democracy Index. *The Economist* attempts to quantify the state of democracy in 167 of the world's countries, focusing on five general categories: electoral processes and pluralism, civil liberties, the functioning of government, political participation, and political culture. The rankings on the democracy index for 2008 were retrieved from http://graphics.eiu.com/PDF/Democracy Index 202008.pdf on June 29, 2009.

Economic freedom. The Heritage Foundation and the *Wall Street Journal* compile this index annually. It covers relative freedoms in ten aspects of the economy (business; trade; property; investments; financial, monetary, and fiscal systems; government; corruption; and labor). The economic freedom index was retrieved from http://www.heritage.org/Index/About.aspx on October 15, 2009.

Days to start business. The goal of the *Doing Business* project was to provide an objective basis for understanding and improving the regulatory environment for business. We used the days required for starting a business as an index of the bureaucratic and legal hurdles an entrepreneur must overcome to incorporate and register a new firm. Data were retrieved from http://doingbusiness.org/ExploreTopics/StartingBusiness/ on June 29, 2009.

Index of shipping difficulties. The index of shipping difficulties is an indicator of the effort required and the complications encountered (border delays, fees, red tape, etc) in the shipping of goods (World Development Report, 2009, Table A4).

Corruption Perception Index. The Corruption Perception Index ranks 180 countries by their perceived levels of freedom from corruption, as determined by expert assessments and public opinion surveys. The index is compiled annually and was retrieved from the Transparency International homepage http://www.transparency.org/policy_research/surveys_indices/cpi on June 29, 2009.

Human Development Index (HDI). The HDI measures the level of human development by combining normalized measures of life expectancy, literacy, educational attainment, and GDP per capita for countries worldwide; the reported indices are for the year 2007 (Human Development Indices, 2008).

GDP. The Gross Domestic Product at Purchasing Power Parity in US Dollars for each country, divided by the midyear population in 2007, was obtained from the Human Development Indices (2008).

Life expectancy. Life expectancy at birth indicates the number of years a newborn infant would live if prevailing patterns of age-specific mortality rates at the time of birth were to stay the same throughout the child's life (Human Development Indices, 2008).

Analytic strategy

Since we are dealing with a sophisticated pattern of relationships – multiple predictors, multiple outcomes, and multiple measures – we will not address each predictor-criterion relationship separately. Instead, we estimate the overlap between respective patterns of relationships. Pattern of relationships means the vector of correlations of one particular facet scale to the whole set of criterion. This way we can formally test how well different facet and rater perspectives (self- vs observer-ratings) agree in their relationships to the criteria and how well the empirical correlations match the predicted correlations. For the formal tests, we calculate correlations between respective vectors of correlations. This is akin to what Westen and Rosenthal (2003) call alerting correlation.

Results

Table 2 reports mean standardized Conscientiousness scores for the two personality databases, NEO PI-R self- and other-ratings. In total, there were data for 60 countries. However, because not all of the data were available for all of the countries, the sample sizes varied across the analyses.

Insert Table 2 about here

The correlations between Conscientiousness scores and the expected criterion variables are shown in Table 3. In order to reduce the possibility of Type I errors resulting from multiple comparisons, a 1% significance threshold was chosen.

The first thing to note about Table 3 is that there are substantially more significant correlations than would be expected by chance. From the 264 correlations in the upper pane (i.e., Pearson moment correlation coefficients before correcting for the effect of GDP), 75 correlations (28%) are significant at $p < .01$. Consistently, a look at the upper pane of Table 3 shows that correlation sizes vary to a great extent but, overall, effect sizes are well above zero. For instance, in the case of aggregate BFI scores, the average of absolute values of correlations with criteria is .34, while for the aggregate NEO PI-R domain scores the respective number is .39. It is worthwhile noting that in individual-level research validity coefficients in this range are interpreted as powerful relationships (e.g. Strenze, 2007). At the level of facets, average effect sizes varied from .10 to .50 and .12 to .40 across facets (respectively for self- and observer-ratings).

Insert Table 3 about here

How well do different facets agree in predicting criteria?

How well do self- and observer-ratings agree in predicting criteria?

Do the observed empirical correlations match predicted correlations?

Self-ratings. The aggregate domain scores of Conscientiousness relate to the criteria similarly, regardless whether they are obtained with the BFI or the NEO PI-R. The correlation between the respective rows (i.e., rows 1 and 2) in Table 3 is .94 ($p < .001$). In some cases, the correlations are consistent with intuitive expectations (though note that we have not formal predictions for domain scores). For example, atheism, alcohol consumption, and the prevalence of smoking tend to relate negatively to average Conscientiousness scores. Countries where people abstain from alcohol and tobacco and are more inclined to believe in God have higher mean Conscientiousness scores. For most of the other criteria, however, correlations are not consistent with intuitive expectations. Health statistics other than alcohol and tobacco use are not significantly correlated to culture-level mean scores of Conscientiousness. Although most societal indicators have significant relationships with Conscientiousness, they run in the opposite direction to expectations. In particular, in countries with higher culture-level mean scores of Conscientiousness, there are more problems with shipping goods and corruption and there is less democracy and economic freedom and lower human development. Or, alternatively, in free, democratic and prosperous countries, people think of themselves as less purposeful, strong-willed, and determined, compared to people in less developed countries.

However the interpretation of the domain scores is complicated because there are clear discrepancies among the different facet scales of Conscientiousness in terms of their relationship to the criteria. In three of the facet scales (C1: Competence, C3: Dutifulness, and C5: Self-Discipline), there are no significant correlations with any criteria. However, the remaining three facet scales (C2: Order, C4: Achievement Striving, and C6: Deliberation) show numerous significant correlations with the criteria and, importantly, they relate to the criteria in a relatively similar manner. The correlations between the respective rows in Table

3 (i.e., rows 4, 6, and 8) range from r = .96 to .97 (p < .001). These three facet scales also match the Conscientiousness domain scores in terms of their relationship to the criteria (in Table 3, the correlations between row 2 and rows 4, 6, and 8 range from .96 to .99).

As a result, it is not surprising that some of the relationships between the facet scales and the criteria were also consistent with our predictions while many of them were not. For instance, aggregate scores on C6: Deliberation had a strong negative relationship with the average amount of alcohol consumed and the prevalence of smoking while it also had a strong negative relationship with democracy and the HDI.

To formally test the degree to which the empirical correlations of the facet scale scores with the criteria matched the predicted correlations, we calculated the correlations between the respective rows in Table 1 and Table 3. It appeared that C2: Order and C4: Achievement Striving generally related to the criteria opposite to predictions (r = -.61 and -.50, with N = 17, coefficients larger than or equal to .49 are significant at p < .05). The third of the facets that had had many significant relationships to the criteria, C6: Deliberation, was more consistent the predictions (r = .15). Interestingly, for C1: Competence and C3: Dutifulness, the pattern of non-significant correlations was in a moderate agreement with theoretical predictions (r = .45 and .38) whereas this was not true for C5: Self-Discipline (r = -.32). Overall, the degree to which predicted correlations matched the empirically obtained correlations varied a lot across different facets of Conscientiousness.

Observer-ratings. Compared to self-ratings, aggregate culture-level observer-ratings seemed to relate to the selected criteria quite differently, at least at first glance. In observer-ratings, there was only one significant correlation at the level of domain scores: Conscientiousness predicted prevalence of smoking (the higher the mean Conscientiousness scores, the more smokers), which is contrary to any standard prediction based on individual-level findings and

also contradicts the relationship found with aggregate self-ratings. To illustrate this reversed pattern more formally, the rows in Table 3 show that the correlations between the aggregate self- and observer-rated Conscientiousness scores and the selected criteria (rows 2 and 9) are significantly negative (r = -.52, p < .05).

However, at the level of facet scales the picture is far more complex. Similarly to the NEO PI-R self-ratings, different facet scales had very different patterns of correlation with the external validity criteria. Two of the facet scales of aggregate observer-rated Conscientiousness (C2: Order and C4: Achievement Striving) had no significant correlations with any criteria. Note that these tended to be the significant predictors in self-ratings. At the same time, those three facet-scales that did not significantly relate to any criteria in case of self-ratings (i.e., C1: Competence, C3: Dutifulness and C5: Self-Discipline), had the strongest correlations in observer-ratings. With the exceptions of smoking and obesity, most of the significant correlations between the aggregate observer-ratings and the external criteria tended to be consistent with the predictions. For instance, higher scores on C1: Competence and C5: Self-Discipline predicted lower mortality due to cardiovascular diseases, injuries, and traffic-accidents, while they were also positively correlated with democracy and human development. Nevertheless, there was a clear exception: aggregate observer-rating scores of the C6: Deliberation correlated with the criteria in a similar way to self-ratings (and thus contrary to predictions). The correlation between the respective rows in Table 3 (rows 8 and 15) was r = .95 (p < .001). Thus, C6: Deliberation behaved differently from the rest of the observer-rated facet scores. To illustrate the differences between facets formally, the profile of correlations for C6: Deliberation (row 15 in Table 3) had strong negative relationships to the correlation profiles of C1: Competence, C3: Dutifulness, and C5: Self-Discipline (rows 10, 12, and 14 in Table 3; r = -.92, -.87, and -.92, respectively, p < .001). Such disagreement

between facets is consistent with our prediction that different facets correlate to the criteria in different manner.

Formal test of the degree to which the empirical correlations match the predicted correlations shows that in three of the facets, C1: Competence, C3: Dutifulness and C5: Self-Discipline, prediction matched empirical results (correlations between the respective rows in Table 1 and Table 3 are r = .53, .47, and .43). These also were the facets with several significant criteria correlations. In C2: Order, the pattern of non-significant correlations tended to be opposite (r = -0.35). In C4: Achievement Striving and C6: Deliberation, there were no systematic trends in the relationships between predicted and empirical correlations (r = .05 and -.07, respectively).

Importantly, at the facet scale level, there was much less contradiction between self- and observer-ratings. The correlations between the respective rows in Table 3—showing how the criteria are correlated to self-ratings on the one side and to observer-ratings on the other —were as follows: r = .48, .69, .39, .09, -.64, and .95, respectively for C1: Competence, C2: Order, C3: Dutifulness, C4: Achievement Striving, C5: Self-Discipline, and C6: Deliberation. The pattern appeared to be contradictory only for C5: Self-Discipline (in which none of the criteria was related to mean self-ratings) and inconsistent for C4: Achievement Striving (in which none of the criteria was related to mean observer-ratings). With respect to the other four facets, the patterns of relationships (although often consisting of non-significant correlations) appear to be moderately to highly similar in terms of shape. Thus, it seems that the contradictory correlations to the criteria are mainly caused by differences in the content of traits (i.e. facet scales) and to a lower degree by differences in the source of ratings (i.e. self- and observer-ratings). It is simply the case that some types of predictor-criterion relationships are stronger from the self-rater perspective while others are stronger from the observer-rater

perspective. In particular, the facets that behaved contrary to expectations were stronger predictors on the self-ratings side, while the facets that behaved expectedly were generally stronger predictors on the observer-ratings side. Thus, what we saw was a highly complex, trait-dependent and observer-perspective moderated pattern of predictor-criterion relationships. This certainly warranted the use of the multi-trait multi-method approach adopted in the current study.

Controlling for National Wealth

We also calculated all correlations taking into account the GDP per capita of the nations (the lower panel of Table 3). For aggregate self-ratings, most of the correlations were substantially reduced after controlling for GDP; only some of the strongest relationships remained significant. After partialling out the effect of GDP, the BFI scores were still significantly negative predictors of economic freedom, human development, and life-expectancy, while the NEO PI-R Conscientiousness domain scores predicted religiosity and C6: Deliberation scores were negatively related to alcohol and tobacco use, ease of starting a business, and human development. Although the correlations became smaller after controlling for GDP, their relative pattern tended to remain similar. In Table 3, the rows showing the correlations between the BFI scores and the criteria before and after controlling for GDP (rows 1 and 16) are highly correlated (r = .95, p < .001). In self-ratings, the relationships between both the NEO PI-R domain and its facet scale scores and the criteria were also similar before and after controlling for the effect of GDP as the correlations between the corresponding rows (rows 2 to 8 and 17 to 23) tend to be moderately to highly positive, ranging from .54 to .96 with a mean of .76. Thus, it appears that the effect of GDP was more or less uniform across all relationships between the aggregate self-rated Conscientiousness scores and the selected criteria.

Partialling out the effect of GDP also tended to lower the correlations between the average observer-rated Conscientiousness scores and the criteria. Yet, more than half of the initially significant correlations remained significant. Again, the effect of controlling for GDP tended to be relatively uniform across all relationships as the corresponding rows in Table 3 are highly correlated. The correlations between rows 9 to 15 on the one side and 24 to 30 on the other (Table 3) range from .60 to .95, with a mean of .86.

Testing formally the degree to which the empirical correlations between Conscientiousness and the selected criteria matched the predicted correlations showed that after controlling for the GDP the pattern of relationships was somewhat more in line with predictions compared to correlations not having GDP taken into account. In self-ratings, correlations between the respective rows of Table 1 and Table 3 (i.e. showing the predicted correlations of Conscientiousness facets and empirical correlations after controlling for the effect of GDP) were r = .22, -.25, .56, .07, -.22, and .32, respectively for C1: Competence, C2: Order, C3: Dutifulness, C4: Achievement Striving, C5: Self-Discipline and C6: Deliberation. Thus, C2: Order and C5: Self- Discipline kept having relationships contrary to prediction, C4: Achievement-Striving was inconsistent in its fit with the predicted correlations to the criteria, and the other three facets agreed modestly to moderately with the predictions. In observer-ratings, five of the six facets agreed with predictions, at least to a moderate extent, after having the GDP being controlled for (r = .38, .20, .45, .33, and .41, for C1: Competence, C2: Order, C3: Dutifulness, C4: Achievement Striving, and C5: Self-Discipline). Only C6: Deliberation, still had a pattern of relationships contrary to predictions (r = -.34).

The overall conclusions are as follows: both aggregate self- and observer-ratings of Conscientiousness had significant relationships with a number of criteria; in some of the facet

scales the relationships tended to be consistent with theoretical predictions, whereas in some facets they tended to be inconsistent with them; thus, different facet scales of Conscientiousness had very different patterns of relationships to the criteria; the facet scales showing significant correlations to the criteria differed across the informant perspective (with the sample size and chosen 0.01 alpha level not too many correlations could be significant); and controlling for variance in national wealth attenuated the relationships but generally improved the agreement between predicted and empirical correlations. Thus, the overall pattern of the relationships of the aggregate personality scores to the selected criteria was extremely complex, though not sporadic.

Discussion

The mass media and the general public alike adore all types of rankings. Indeed, it would be worthwhile to know where on this planet the most purposeful, strong-willed, and determined people live. However, Heine and his colleagues (2008) recently issued a warning that people's ratings of their own personality, at least, cannot be trusted as a reliable foundation for this kind of ranking.

In this paper we want to argue that the actual job of investigating the validity of culture-level aggregate personality scores has only begun. To date there is very few evidence for any solid conclusion on the validity of the aggregate scores. Conscientiousness is the personality trait that has received somewhat more attention than the other traits, as it has been shown to relate to external criteria in a manner not consistent with researchers' predictions. However, even for the Conscientiousness, there is far too little rigorous research to be able to come up with any definitive answers.

In his paper we extended previous research on the predictive validity of Conscientiousness in three important ways. Firstly, we used a more thorough and

theoretically organized set of criterion for culture-level personality scores. Secondly, we set up precise hypotheses about the predictor-criterion relationships, thereby making the hypotheses testing more transparent and formal. Thirdly, we broke Conscientiousness into lower-level facets assuming that different sub-themes of the broad trait may relate to criteria in a different manner. All of these extensions proved worthwhile as the new criteria were often related to personality scores, setting up precise numerical hypotheses and comparing those to the empirical findings helped to organize the complex pattern of findings and reveal regularities in the patterns, and different facets of Conscientiousness appeared to relate to the criteria in a very different manner.

One of the most important messages of this study is that there are no predetermined criterion variables, even if many of them may look promising. As was discussed in the Introduction, ideally, the way in which validity-criteria are chosen should be theoretically guided and the relationships involved should be explicitly described. For instance, if we hypothesize that higher mean culture-level scores of Conscientiousness are related to higher levels of economic wealth, as reflected in high GDP, then we should be able to propose the exact mechanism of how a high proportion of purposeful, strong-willed, and determined people leads to the economic success of the whole country. At present, however, researchers (including us) have only relatively vague speculations about this. Without clear and empirically supported causal explanations, relationships can be rather unpredictable. In such a situation, it is possible to rely on multiple criteria and look at the pattern of relationships to sort out meaningful ones.

Consistently with previous studies we showed that Conscientiousness scores have many relationships to the external criteria that are not consistent with intuitive expectations. However, the use of facet scale scores in addition to Conscientiousness domain scores

allowed us to show that the associations of Conscientiousness with the criteria were actually much more complex. Not all but only some of aspects of Conscientiousness—Order, Achievement Striving and Deliberation—tended to relate to criteria in the manner, which was contrary to predictions (especially in self-ratings). For Competence and Dutifulness, the pattern of relationships to the criteria tended to be consistent with our prediction, although the correlations were non-significant in self-ratings. For Self-Discipline, the patterns were contradictory in self- and observer-ratings.

It is interesting that expected correlations tended to be stronger in observer-ratings compared to self ratings, indicating that aggregate observer-rating scores may in many cases be more veridical indicators of cross-cultural personality differences. However, it is important to note that this is not an absolute rule as in Deliberation, self-rating scores tended to be more in line with expectations compared to observer-ratings. As a result, it is not readily possible to say that there are some mechanisms in self-ratings that is not present in observer ratings and that makes aggregate scores invalid. One of the explanations why self- and observer-ratings differed is related to sampling. Self-ratings described mainly students, while observer-ratings also described older people. It is possible, for instance, that some of the aspects of Conscientiousness are more consequential in terms of real-life outcomes (or, alternatively, are influenced by culture-level variables) in younger age while the other aspects are more important (or more influenced) in older age.

What this facet-dependent and observer-perspective moderated relationship tells us is that, on the basis of existing data, there is no simple answer for the question of the validity of culture-level personality scores. These findings do not necessarily support the positions that there is a general X-factor such as reference standard effect (Heine et al., 2008) that twists the culture-level personality scores and makes them incomparable. If there is a reference standard

effect, it has to be specific to (some) facets of Conscientiousness. Nor do these findings necessarily support the opposite position saying that culture-level Conscientiousness scores are fine and can be used without any worries about their validity. Nor can we say that self-ratings do not work but other-ratings do: aggregate observer-rated Deliberation related to the criteria in a manner, which was not consistent with predictions. We realize that readers would prefer a clear answer to the question of validity of culture-level Conscientiousness but there simply is none. We hope that this article helped to show that the problem cannot be solved calculating the relationships of aggregate personality scores to a couple of seemingly likely culture-level criterion variables. The problem is more complex.

In this study, we laid down a more rigorous agenda for studying the criterion-related validity of culture-level aggregate personality scores and we also documented a complex pattern of relationships that we are not as yet able to make full sense of. Some of the findings were consistent with our predictions and thereby speak for the validity of culture-level personality scores. Some of the findings did not. The findings that were contrary to predictions may indicate that in some facets culture-level personality scores are indeed not valid. Importantly, however, they may also be the result of the wrong choice of criteria or incorrectly formulated hypotheses. Thus, there is urgent need for more rigorous and theoretically refined research on the validity of culture-level personality scores.

References

Adewuya, A. O., Loto, O. M., & Adewumi, T. A. (2008). Premenstrual dysphoric disorder amongst Nigerian university students: prevalence, comorbid conditions, and correlates. *Archives of Women's Mental Health, 11*, 13-18.

Allik, J., & McCrae, R. R. (2004). Toward a geography of personality traits - Patterns of profiles across 36 cultures. *Journal of Cross-Cultural Psychology, 35*, 13-28.

Ashton, M. C. (2007). Self-reports and stereotypes: A comment on McCrae et al. *European Journal of Personality, 21*, 983-986.

Aziz, S., & Jackson, C. J. (2001). A comparison between three and five factor models of Pakistani personality data. *Personality and Individual Differences, 31*, 1311-1319.

Barrick, M. R., & Mount, M. K. (1991). The big five personality dimensions and job performance: a meta-analysis. *Personnel Psychology, 44*, 1-26.

Benet-Martinez, V., & John, O. P. (1998). Los Cinco Grandes across cultures and ethnic groups: Multitrait multimethod analyses of the Big Five in Spanish and English. *Journal of Personality and Social Psychology, 75*, 729-750.

Bray, I., & Gunnell, D. (2006). Suicide rates, life satisfaction and happiness as markers for population mental health. *Social Psychiatry and Psychiatric Epidemiology, 41*, 333-337.

Connelly, B. S., & Ones, D. S. (2008). The Personality of Corruption A National-Level Analysis. *Cross-Cultural Research, 42*, 353-385.

Costa, P. T., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual*. Odessa, Fl.: Psychological Assessment Resources.

Deary, I. J., Batty, D. G., Pattie, A., & Gale, C. R. (2008). More Intelligent, More Dependable Children Live Longer: A 55-Year Longitudinal Study of a Representative Sample of the Scottish Nation. *Psychological Science, 19*, 874-880.

Diener, E., & Diener, C. (1995). The wealth of nations revisited: Income and quality of life. *Social Indicators Research, 36*, 275-286.

Friedman, H. S., Tucker, J. S., Schwartz, J. E., Tomlinson-Keasey, C., & Martin, L. R. (1995). Psychosocial and behavioral predictors of longevity. *American Psychologist, 50*, 69-78.

Heine, S. J., Buchtel, E. E., & Norenzayan, A. (2008). What do cross-national comparisons of personality traits tell us? The case of conscientiousness. *Psychological Science, 19*, 309-313.

Hofstede, G., & McCrae, R. R. (2004). Personality and culture revisited: linking traits and dimensions of culture. *Cross-Cultural Research, 38*, 52-88.

Human Development Index 2007 (2009). *Human Development Report.* New York: The United Nations Development Program.

Inglehart, R. (1990). *Culture shift in advanced industrial society*. Princeton, NJ: Princeton University Press.

Inglehart, R., & Baker, W. E. (2000). Modernization, cultural change, and the persistence of traditional values. *American Sociological Review, 65*, 19-51.

Inglehart, R., & Welzel, C. (2005). *Modernization, cultural change, and democracy: The human development sequence*. New York, NY: Cambridge University Press.

Kern, M. L., & Friedman, H. S. (2008). Do conscientious individuals live longer? A quantitative review. *Health Psychology, 27*, 505-512.

Koivumaa-Honkanen, H., Honkanen, R., Viinamaki, H., Heikkila, K., Kaprio, J., & Koskenvuo, M. (2001). Life satisfaction and suicide: A 20-year follow-up study. *American Journal of Psychiatry, 158*, 433-439.

Lodi-Smith, J., & Roberts, B. W. (2007). Social investment and personality: A meta-analysis of the relationship of personality traits to investment in work, family, religion, and volunteerism. *Personality and Social Psychology Review, 11*, 68-86.

Lucas, R. E., Diener, E., Grob, A., Suh, E. M., & Shao, L. (2000). Cross-Cultural Evidence for the Fundamental Features of Extraversion. *Journal of Personality and Social Psychology, 79*, 452-468.

Lönnqvist, J.-E., Paunonen, S., Tuulio-Henriksson, A., Lönnqvist, J., & Verkasalo, M. (2007). Substance and style in socially desirable responding. *Journal of Personality, 75*, 291-322.

Lynn, R., Harvey, J., & Nyborg, H. (2009). Average intelligence predicts atheism rates across 137 nations. *Intelligence, 37*, 11-15.

McCrae, R. R. (2002). NEO-PI-R data from 36 cultures: further intercultural comparisons. In R. R. McCrae & J. Allik (Eds.), *The Five-Factor Model of personality across cultures* (pp. 105-125). New York: Kluwer Academic/Plenum Publishers.

McCrae, R. R., & Terracciano, A. (2008). The Five-Factor Model and its correlates in individuals and cultures. In F. van de Vijver, D. A. van Hemert & Y. H. Poortinga (Eds.), *Multilevel analysis of individuals and cultures* (pp. 249-283). New York: Lawrence Erlbaum Associates.

McCrae, R. R., Terracciano, A., & 78 Members of the Personality Profiles of Cultures Project. (2005). Universal features of personality traits from the observer's

perspective: Data from 50 cultures. *Journal of Personality and Social Psychology, 88*, 547-561.

McCrae, R. R., Terracciano, A., & 79 Members of the Personality Profiles of Cultures Project. (2005). Personality Profiles of Cultures: Aggregate Personality Traits. *Journal of Personality and Social Psychology, 89*, 407-425.

McCrae, R. R., Terracciano, A., Realo, A., & Allik, J. (2007). On the validity of culture-level personality and stereotype scores. *European Journal of Personality, 21*, 987-991.

McCrae, R. R., Terracciano, A., Realo, A., & Allik, J. (2008). Interpreting GLOBE Societal Practices Scales. *Journal of Cross-Cultural Psychology, 39*, 805-810.

Moon, H. (2001). The two faces of conscientiousness: duty and achievement striving in escalation of commitment dilemmas. *The Journal of Applied Psychology*, *86*(3), 533-540.

Oishi, S., & Roth, D. P. (2009). The role of self-reports in culture and personality research: It is too early to give up on self-reports. *Journal of Research in Personality, 43*, 107-109.

Ozer, D. J., & Benet-Martinez, V. (2006). Personality and the prediction of consequential outcomes. *Annual Review of Psychology, 57*, 401-421.

Perugini, M., & Richetin, J. (2007). In the land of the blind, the one-eyed man is king. *European Journal of Personality, 21*, 977-981.

Rentfrow, P. J., Gosling, S. D., & Potter, J. (2008). A Theory of the Emergence, Persistence, and Expression of Geographic Variation in Psychological Characteristics. *Perspectives on Psychological Science, 3*, 339-369.

Roberts, B. W., Chernyshenko, O. S., Stark, S., & Goldberg, L. R. (2005). The structure of conscientiousness: an empirical investigation based on seven major personality questionnaires. *Personnel Psychology, 59*, 103-139.

Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review, 15*, 351-357.

Saroglou, V. (2010). Religiousness as a Cultural Adaptation of Basic Traits: A Five-Factor Model Perspective. *Personality and Social Psychology Review, 14*, 108-125.

Schmidt, F. L., & Hunter, J. (2004). General mental ability in the world of work: occupational attainment and job performance. *Journal of Personality and Social Psychology, 86*, 162-173.

Schmitt, D. P., Allik, J., McCrae, R. R., & Benet-Martinez, V. (2007). The geographic distribution of big five personality traits - Patterns and profiles of human self-description across 56 nations. *Journal of Cross-Cultural Psychology, 38*, 173-212.

Siuta, J. (2006). *Inwentarz osobowosci NEO-PI-R Paula T.Costy Jr. i Roberta R.McCrae. Adaptacja polska. Podrecznik.[Personality inventory NEO-PI-R by Paul T.Costa, Jr. and Robert R.McCrae. Polish adaptation. A handbook]*. Warszawa: Pracownia Testów Psychologicznych.

Strenze, T. (2007). Intelligence and socioeconomic success: A meta-analytic review of longitudinal research. *Intelligence, 35,* 401-426.

Zhao, H., & Seibert, S. E. (2006). The Big Five personality dimensions and entrepreneurial status: A meta-analytical review. *Journal of Applied Psychology, 91,* 259-271.

Zuckerman, P. (2007). Atheism: contemporary numbers and patterns. In M. Martin (Ed.), *The Cambridge companion to atheism* (pp. 47-65). Cambridge: Cambridge University Press.

Zukauskiene, R., & Barkauskiene, R. (2006). Lietuviškosios NEO PI-R versijos psichometriniai rodikliai [Psychometric properties of the Lithuanian version of the NEO PI-R]. *Psichologija, 33,* 7-21.

Terracciano, A., Sutin, A. R., McCrae, R. R., Deiana, B., Ferrucci, L., Schlessinger, D., et al. (in press). Facets of Personality Linked to Underweight and Overweight. *Psychosomatic Medicine.*

Westen, D., & Rosenthal, R. (2003). Quantifying construct validity: Two simple measures. *Journal of Personality and Social Psychology, 84*(3), 608-618.

Williams, E. B. (1979). *The new Bantam English disctionary*. New York: Bantam Books.

World Development Report. (2009). *Reshaping economic geography*. Washington, DC: The World Bank.

World Health Organization. (2009). World Health Statistics 2009. *WHO Statistical Information System (WHOSIS).* The data was retrieved from http://www.who.int/whosis/en/index.html on January 20, 2010.

Author Notes

This project was supported by grants from the Estonian Ministry of Science and Education (SF0180029s08) and the Estonian Science Foundation (ESF7020) to Jüri Allik. We are very thankful to Delaney Michael Skerrett for his helpful comments and suggestions on an earlier draft of this manuscript.

Table 1. *Predicted Correlations between the Conscientiousness Facets of the NEO PI-R and Chosen External Criteria*

	Atheism	Alcohol consumption	Smoking	Obesity (males)	Obesity (females)	HIV	Cardiovascular mortality	Injury-related mortality	Traffic deaths	Homicides	Democracy	Economic Freedom	Starting Business	Shipping Difficulties	Corruption	HDI	GDP	Life-Expectancy
C1:Competence	-	-	-	-	-	-	-	-	-	-	++	+	-	-	0	++	++	++
C2:Order	--	-	-	-	-	-	-	-	-	-	-	++	--	--	0	+	+	+
C3:Dutifulness	--	-	-	-	-	-	-	-	-	-	0	++	--	--	--	+	+	+
C4:Achievement Striving	--	-	-	-	-	-	-	-	-	-	-	+	-	-	0	+	+	+
C5:Self-Discipline	--	--	--	--	--	--	--	--	--	--	0	0	0	0	-	+	+	+
C6:Deliberation	0	--	--	--	--	--	--	--	--	--	--	--	++	++	--	0	0	+

Note: ++ predicted correlations in the moderate to strong size (.50); + predicted correlations in the small to moderate size (.30); - predicted correlations in the small to moderate size (-.30); -- predicted correlations in the moderate to strong size (-.50). According to Costa & McCrae (1992), C1: Competence "refers to the sense that one is capable, sensible, prudent, and effective", C2: Order refers to the degree to which people are neat, tidy and well organized, C3: Dutifulness refers to being governed by conscience, C4: Achievement Striving refers to degree to which people are purposeful and work hard, C5: Self-Discipline refers to "the ability to begin tasks and carry them through to completion despite boredom and other distractions", and C6: Deliberation refers to "the tendency to think carefully before acting."

Table 2. *Mean National Scores of Conscientiousness.*

	Self-ratings							Observer-ratings						
Country	C	C1	C2	C3	C4	C5	C6	C	C1	C2	C3	C4	C5	C6
Argentina								50.0	50.6	47.6	51.1	52.6	52.0	48.0
Australia								47.5	51.0	47.6	48.4	47.9	49.4	47.5
Austria	46.7	47.2	47.4	46.7	49.4	43.7	47.6	52.4	54.5	52.8	52.9	50.0	53.2	50.5
Belgium	46.6	43.8	47.2	48.7	48.5	46.1	49.1	47.4	47.4	47.7	48.6	49.6	49.0	48.4
Brazil								51.5	51.6	50.7	51.8	51.5	50.1	49.7
Canada	49.2	50.7	49.3	49.8	47.5	48.3	50.8	49.6	50.7	50.3	49.9	48.8	50.5	50.9
Chile								52.2	54.1	49.9	52.8	53.6	52.6	50.4
China	50.3	44.0	47.7	50.5	49.7	47.2	57.2	48.0	48.3	48.9	48.0	47.4	48.6	52.0
Congo	52.8	47.1	53.9	50.0	55.6	47.7	58.4							
Congo (Dem Rep.)	52.8	50.1	53.5	48.4	54.4	48.1	59.5							
Croatia	53.2	47.6	50.2	51.5	54.6	48.5	51.8	50.3	50.6	50.5	49.5	52.4	49.7	48.3
Czech Rep.	47.5	40.3	47.7	50.1	49.7	45.2	49.8	51.5	49.0	48.3	53.9	50.7	50.3	50.5
Denmark	47.5	48.2	48.7	50.6	48.5	49.4	48.5	48.4	53.1	47.4	50.4	47.9	48.6	47.9
Estonia	49.6	44.6	50.3	52.4	50.2	49.6	50.6	50.0	49.7	51.4	49.7	50.0	50.6	48.9
Ethiopia								47.2	46.0	50.8	44.3	47.7	47.7	51.8
Finland	48.3	50.3	51.5	49.3	49.3	46.5	46.0							
France	47.4	42.1	48.3	49.2	48.2	44.7	48.0	48.4	48.0	48.3	48.9	47.0	48.4	47.1
Germany	46.7	45.3	48.7	46.7	48.2	44.6	47.0	52.3	54.3	51.8	53.6	50.4	52.3	50.5
Hong Kong	49.2	40.3	48.6	48.6	48.7	48.7	53.4	49.6	46.1	51.2	49.6	48.5	49.1	50.3
Hungary	50.0	42.5	51.8	51.2	50.3	45.8	49.5							
Iceland								49.3	50.2	49.5	51.5	51.6	49.7	50.5
India	54.9	45.8	54.1	53.1	54.3	48.9	55.9	52.3	47.8	53.9	51.3	53.7	50.2	52.8
Indonesia	50.3	42.0	52.1	49.2	54.3	45.9	56.4	49.6	45.5	51.3	48.4	50.2	47.8	52.4
Iran								47.0	47.4	48.7	47.3	49.4	45.9	48.5
Italy	50.4	44.1	45.0	51.3	49.3	48.2	51.9	48.3	47.0	47.2	48.2	47.9	48.8	48.1
Japan	42.6	34.9	45.6	43.2	45.9	39.8	48.0	49.5	48.0	48.6	48.9	50.6	48.9	47.9
Korea, Rep of	48.8	42.1	47.6	52.8	47.9	44.9	52.5	48.3	48.4	49.1	49.3	45.7	50.7	50.4
Kuwait								52.6	51.0	50.5	51.9	52.9	50.7	49.9
Lebanon								50.5	48.3	50.7	49.1	49.4	50.5	50.8
Lithuania	46.1	41.3	48.9	49.0	47.5	44.8	50.1							
Malaysia	54.2	44.9	56.3	53.2	55.0	44.5	56.9	53.0	49.8	53.1	52.1	53.1	48.6	54.6
Malta								51.6	51.3	50.1	50.9	48.9	51.3	49.8
Mauritius	49.2	45.2	50.7	49.3	51.8	46.9	52.6							
Mexico								50.7	51.3	49.6	50.3	52.0	51.7	52.0
Morocco								45.5	43.9	47.6	43.8	44.9	45.9	48.2
Netherlands	48.6	45.6	48.7	52.5	49.6	50.0	51.2							
New Zealand								47.8	51.9	48.0	48.5	48.6	50.3	48.0
Nigeria								45.8	45.0	47.7	44.2	47.0	46.3	50.2
Norway	45.7	47.8	48.2	49.0	48.8	45.8	48.6							
Peru	49.0	47.8	46.5	49.4	53.1	44.8	50.9	48.7	49.4	48.3	48.6	50.5	49.3	49.1
Philippines	51.5	47.1	50.7	49.0	52.4	49.4	54.7	53.5	53.0	51.4	51.1	54.0	52.3	51.8

Table 2. *Mean National Scores of Conscientiousness (continued).*

	Self-ratings							Observer-ratings						
Country	C	C1	C2	C3	C4	C5	C6	C	C1	C2	C3	C4	C5	C6
Poland	47.9	42.9	51.5	49.5	50.7	47.1	48.0	49.4	48.3	51.5	48.4	49.9	48.5	47.9
Portugal	50.3	44.9	50.8	50.2	51.7	46.4	51.4	50.7	51.5	51.5	51.1	51.4	50.2	49.9
Puerto Rico								52.9	53.1	49.2	52.1	53.4	52.8	51.8
Russia	46.5	40.6	50.2	46.0	46.8	43.8	50.3	49.1	47.1	51.0	49.3	48.3	47.3	51.6
Serbia	51.7	46.9	50.3	51.3	54.2	47.7	51.0	51.7	52.8	50.1	51.9	52.2	51.6	50.3
Slovakia								48.6	49.7	50.2	49.0	47.8	48.4	51.0
Slovenia								52.3	51.4	51.5	51.3	52.0	51.2	52.1
South Africa	47.9	44.8	47.9	47.1	47.7	47.1	51.8							
Spain	48.3	44.6	48.1	47.8	51.4	44.3	51.8	51.3	52.7	51.2	51.7	51.3	52.1	51.3
Sweden	45.7	48.8	49.8	52.7	42.7	47.0	54.5							
Switzerland	49.6	48.9	50.6	48.6	50.6	45.6	48.7	51.6	52.2	50.7	52.7	49.9	52.0	49.4
Taiwan	48.1	42.5	47.3	48.9	49.7	45.7	54.4							
Thailand								48.9	49.5	48.8	49.4	49.2	48.8	51.2
Turkey	50.4	49.5	47.3	50.2	52.0	48.2	51.4	51.4	50.9	51.6	52.0	51.1	49.9	50.2
Uganda								48.2	47.4	50.2	44.8	48.5	48.3	51.5
United Kingdom								48.1	51.3	49.0	48.8	48.6	49.4	47.8
United States	50.0	50.0	50.0	50.0	50.0	50.0	50.0	48.8	51.0	49.4	48.5	49.5	50.8	49.0
Zimbabwe	51.8	40.9	53.1	49.3	55.3	48.1	55.2							

Note. BFI = Big Five Inventory (Benet-Martinez & John, 1998); NEO PI-R = Revised NEO Personality Inventory (Costa & McCrae, 1992); C = Conscientiousness; C1 = Competence; C2 = Order; C3 = Dutifulness; C4 = Achievement Striving; C5 = Self-Discipline; C6 = Deliberation.

Table 3**.** *Correlations between Mean National Scores of Conscientiousness and External Criterion Variables.*

		Atheism	Alcohol consumption	Smoking	Obesity (males)	Obesity (females)	HIV	Cardiovascular mortality	Injury-related mortality	Traffic deaths	Homicides	Democracy	Economic Freedom	Starting Business	Shipping Difficulties	Corruption	HDI	GDP	Life-Expectancy
	Pearson Product Moment Correlations																		
	SELF RATINGS BFI (*N*=38-53)																		
1.	Conscientiousness	**-.44**	-.26	**-.42**	.20	-.11	.37	.26	**.38**	.33	.17	**-.43**	**-.41**	.22	**.47**	-.31	**-.53**	-.33	**-.52**
	SELF RATINGS NEO-PI-R (*N*=32-39)																		
2.	Conscientiousness	**-.67**	**-.49**	-.34	-	-.06	.13	.33	.22	.34	-.05	**-.50**	**-.48**	.34	**.43**	**-.54**	**-.59**	**-.68**	**-.42**
3.	C1:Competence	-.22	-.17	-.26	-	.12	-.13	-.15	-.22	-.26	-.06	.14	.10	.03	.01	.20	-.15	.05	.00
4.	C2:Order	-.39	-.41	-.33	-	-.04	.13	**.43**	.40	.30	-.02	-.38	**-.42**	.32	.34	-.37	**-.52**	**-.50**	**-.44**
5.	C3:Dutifulness	-.12	-.18	-.06	-	.02	-.17	-.01	-.17	-.03	-.25	.07	-.03	-.19	-.15	-.03	-.04	-.15	.04
6.	C4:Achievement Striving	**-.68**	**-.49**	-.24	-	-.17	.16	.32	.27	.33	-.11	**-.55**	**-.57**	**.50**	**.46**	**-.63**	**-.59**	**-.61**	**-.50**
7.	C5:Self-Discipline	-.24	-.21	-.29	-	.17	.15	.18	.02	.04	.04	-.12	-.05	.02	.13	-.07	-.22	-.11	-.23
8.	C6:Deliberation	**-.47**	**-.72**	**-.54**	-	-.35	.23	.38	**.46**	**.48**	.08	**-.67**	**-.55**	**.57**	**.48**	**-.62**	**-.74**	**-.64**	**-.60**
	OBSERVER RATINGS NEO-PI-R (*N*=34-47)																		
9.	Conscientiousness	-.04	-.07	**.43**	.07	.20	-.29	-.10	-.18	-.30	.14	.25	.03	.18	-.17	.06	.25	.12	.28
10.	C1:Competence	.14	.36	**.49**	**.55**	**.53**	-.34	**-.49**	**-.54**	**-.57**	.08	**.54**	**.38**	.01	-.35	**.45**	**.53**	.38	**.49**
11.	C2:Order	-.21	-.24	.23	-.06	-.03	-.00	.29	.22	.05	.06	.00	-.11	.16	-.03	-.09	-.12	-.10	-.09
12.	C3:Dutifulness	.21	.22	**.61**	.12	.32	**-.51**	-.31	-.37	**-.55**	.06	**.49**	.25	.06	**-.41**	.34	**.58**	.39	**.56**
13.	C4:Achievement Striving	-.19	-.19	.27	.01	.20	-.19	-.08	-.11	-.11	.18	.13	-.10	.25	.03	-.09	.10	-.05	.13
14.	C5:Self-Discipline	.07	.25	.38	**.47**	.41	-.30	**-.48**	**-.43**	**-.54**	.06	**.53**	.35	-.03	-.33	.34	**.46**	.35	**.47**
15.	C6:Deliberation	-.33	-.41	-.11	**-.25**	-.22	.20	**.30**	**.41**	**.35**	.20	**-.25**	-.26	.15	.20	-.35	**-.40**	-.39	**-.37**

Table 3**.** *Correlations between Mean National Scores of Conscientiousness and External Criterion Variables (continued).*

		Atheism	Alcohol consumption	Smoking	Obesity (males)	Obesity (females)	HIV	Cardiovascular mortality	Injury-related mortality	Traffic deaths	Homicides	Democracy	Economic Freedom	Starting Business	Shipping Difficulties	Corruption	HDI	GDP	Life-Expectancy
	Partial Correlations (The effect of GDP is partialed out)																		
	SELF RATINGS BFI (*N*=38-53)																		
16.	Conscientiousness	-.32	-.11	-.34	.34	-.03	.30	.06	.24	.17	.08	-.30	-.26	.10	.36	-.06	**-.48**	-	**-.52**
	SELF RATINGS NEO-PI-R (*N*=32-39)																		
17.	Conscientiousness	**-.49**	-.27	-.17	-	.11	-.04	-.11	-.18	-.10	-.31	-.18	-.01	.13	.08	-.12	-.24	-	-.05
18.	C1:Competence	-.31	-.23	-.31	-	.11	-.12	-.16	-.24	-.30	-.05	.14	.06	.06	.05	.29	-.33	-	-.06
19.	C2:Order	-.19	-.24	-.19	-	.07	.00	.20	.20	.00	-.19	-.12	.13	.17	.09	.00	-.32	-	-.29
20.	C3:Dutifulness	-.04	-.12	-.00	-	.06	-.23	-.15	-.32	-.19	-.32	.24	.05	-.28	-.33	.19	.15	-	.24
21.	C4:Achievement Striving	**-.50**	-.25	-.01	-	-.02	-.02	-.17	-.15	-.16	**-.42**	-.23	-.09	.34	.10	-.27	-.20	-	-.08
22.	C5:Self-Discipline	-.22	-.18	-.26	-	.21	.13	.15	-.05	-.05	.00	-.07	.16	-.02	.07	.05	-.23	-	-.23
23.	C6:Deliberation	-.14	**-.60**	**-.44**	-	-.26	.07	-.09	.14	.08	-.18	-.41	.19	**.44**	.12	-.19	**-.49**	-	-.23

Table 3**.** *Correlations between Mean National Scores of Conscientiousness and External Criterion Variables (continued).*

		Atheism	Alcohol consumption	Smoking	Obesity (males)	Obesity (females)	HIV	Cardiovascular mortality	Injury-related mortality	Traffic deaths	Homicides	Democracy	Economic Freedom	Starting Business	Shipping Difficulties	Corruption	HDI	GDP	Life-Expectancy
	OBSERVER RATINGS NEO-PI-R (*N*=34-47)																		
24.	Conscientiousness	-.14	-.17	**.45**	.04	.18	-.26	-.03	-.13	**-.30**	.19	.24	.14	.25	-.12	-.09	.27	-	**.40**
25.	C1:Competence	-.12	.20	**.42**	**.49**	**.48**	-.26	-.34	**-.42**	**-.46**	.24	**.42**	-.01	.19	-.14	.26	.38	-	.33
26.	C2:Order	-.19	-.22	.32	-.03	-.01	-.03	.31	.20	-.02	.02	.10	.32	.14	-.13	-.01	-.07	-	-.03
27.	C3:Dutifulness	-.03	.02	**.56**	.00	.25	**-.46**	-.07	-.19	**-.42**	.22	.34	.15	.26	-.22	.03	**.49**	-	**.44**
28.	C4:Achievement Striving	-.21	-.19	.34	.03	.22	-.22	-.15	-.17	-.20	.18	.23	.03	.25	.00	-.10	.25	-	.25
29.	C5:Self-Discipline	-.18	.09	.31	.40	.35	-.23	-.35	-.30	**-.44**	.19	**.42**	.13	.14	-.15	.10	.33	-	.34
30.	C6:Deliberation	-.14	-.26	.05	-.15	-.14	.10	.06	.25	.12	.08	.02	.31	-.01	-.07	-.04	-.16	-	-.14

Note: Significant correlations p < .01 are shown in boldface; Correlations that remain or become significant after the impact of GDP was partialed out are underlined.